

Amendments to the Drawings:

Subject to the approval of the Examiner, please replace the drawing sheet labeled Fig. 8 with the attached Replacement Drawing Sheet Fig. 8. The Replacement Drawing Sheet Fig. 8 has been amended to include the words "Yes" and "No" after the decision block 802.

Remarks

In the Office Action, the Examiner rejects claims 1-8 and 31 under 35 U.S.C. § 112, second paragraph, as indefinite; rejects claims 1, 2, and 9-31 under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 6,665,658 to DaCosta et al. ("DaCosta") in view of the publication "Logic Programming with the World-Wide Web," by Loke et al. ("Loke"); and rejects claims 3-8 under 35 U.S.C. § 103(a) as being unpatentable over DaCosta and Loke, and further in view of web document "Verity, Ultraseek, Support FAQ #1037" ("Ultraseek").

By this Amendment, Applicants amend claims 1, 3, 4, 9, and 31 to more appropriately define the invention. Additionally, Applicants have amended the specification to correct obvious typographical errors relating to reference numerals in the drawings and propose amending Fig. 8 to add the words "Yes" and "No" after the decision block 802.

Claims 1-31 remain pending.

Claims 1-8 and 31 stand rejected under 35 U.S.C. § 112, second paragraph, because, according to the Examiner, the phrase "based, at least in part" renders these claims indefinite. Although not agreeing with the Examiner, in order to expedite prosecution, Applicants have amended claims 1, 3 and 31 to delete "at least in part." In view of these amendments, Applicants submit that the rejections of claims 1-8 and 31 under 35 U.S.C. § 112, second paragraph, are obviated.

*Rejections Under 35 U.S.C.
§ 103(a) Based on DaCosta and Loke*

Claims 1, 2, and 9-31 stand rejected under 35 U.S.C. § 103(a) based on DaCosta and Loke. For the following reasons, Applicants respectfully disagree with the rejections of these claims.

Claim 1, as amended, is directed to a method for crawling documents comprising receiving a uniform resource locator (URL) and receiving at least two different copies of a document associated with the URL. The method further includes determining whether a web site corresponding to the URL uses session identifiers based on a comparison of URLs that are within the document and that change between the at least two different copies of the document. DaCosta and Loke, either alone or in combination, do not disclose or suggest these features of claim 1.

DaCosta, in contrast to claim 1, is directed to a web crawler to automatically simulate user interaction with a dynamic website in order to gather and extract information from the site. (DaCosta, Abstract). To this end, DaCosta discloses determining whether a URL corresponds to a dynamic web site. (DaCosta, column 4, lines 41-46). Determining whether a website is dynamic, however, is not the same, and does not disclose or suggest determining whether a web site uses session identifiers.

A “session identifier,” as is known in the art and is consistently used in the pending specification, refers to embedded information within the URL of a web page. (See Spec., paragraphs 0006, 0042, and 0043; and Fig. 5). Session

identifiers are commonly used by web sites to track user behavior as they traverse a web site.

The Examiner contends that DaCosta discloses “a method for crawling documents in a dynamic website, with a database for storing and identifying session identifier URLs.” (Office Action, page 3). DaCosta discloses a web crawler that simulates user interaction with a dynamic web site. (DaCosta, Abstract). As is known in the art and as described by DaCosta, a dynamic web site is one that generates content dynamically in response to user interaction. (DaCosta, col. 1, lines 18-45). DaCosta, however, does not mention the use of session identifiers, much less the specific acts recited in amended claim 1, including, for example, “determining whether a web site corresponding to the URL uses session identifiers based on a comparison of URLs that are within the document and that change between the at least two different copies of the document.”

The Examiner points to portions of columns 4, 5, and 6 of DaCosta as disclosing the use of session identifiers. Specifically, the Examiner appears to contend that DaCosta, at column 4, lines 41 through column 5, line 23 and column 6, lines 21-40, discloses session identifiers. (Office Action, page 3). Applicants respectfully disagree with the Examiner’s interpretation of DaCosta.

Column 4, line 31 through column 5, line 23 of DaCosta discusses, among other things, “session data” of a web site. For example, DaCosta states:

It is also preferred that the step of determining if said URL is a dynamic website further comprise performing a hypertext transfer protocol GET method of the website, downloading a content including a header of the website, and scanning the header for the session data which may be represented by a cookie.

(DaCosta, column 4, lines 40-46) (emphasis added). The “session data” discussed in this section of DaCosta appears to broadly relate to any session information used in the context of dynamic generation of web content. DaCosta discloses that the session data may be obtained by scanning a header of a website, and that the session data may be represented by a cookie. The Examiner can appreciate session data represented by a cookie, as disclosed by DaCosta, cannot be said to be equivalent to the session identifier recited in claim 1. As described in the pending specification, a cookie, although may be used to track user behavior, is different than a session identifier. (Spec., first sentence of paragraph 0006).

The Examiner also points to column 6, lines 21-40 of DaCosta as disclosing the session identifiers recited in claim 1. This section of DaCosta, however, again discusses “session data” that may be represented in a cookie. DaCosta, as discussed above, does not disclose using session identifiers. DaCosta therefore, could not possibly disclose or suggest, as is recited in claim 1, determining whether a web site corresponding to a URL uses session identifiers.

Because DaCosta does not disclose determining whether a web site uses session identifiers, DaCosta could not possibly disclose or suggest, as is also recited in claim 1, determining whether a web site uses session identifiers based on a comparison of URLs that are within the document and that change between the at least two different copies of the document.

Amended claim 1 also recites “receiving at least two different copies of a document associated with the URL.” DaCosta in no way discloses or suggests this feature of claim 1.

Loke does not cure the above-noted deficiencies of DaCosta. Although Applicants are not clear as to which features of claim 1 the Examiner is relying upon Loke as disclosing or suggesting, Applicants submit that Loke, as with DaCosta, does not disclose or suggest making any kind of determination as to whether a web site uses session identifiers, much less the specific techniques for making this determination that are recited in claim 1.

The Examiner states that “Loke teaches the use of URLs with session identifiers which can be extracted when required (p. 4, par. 47-57).” (Office Action, page 3). Applicants note that Loke extends from pages 235 to 245 and that the Examiner’s citing of page 4 of Loke appears to be an error. In any event, Applicants submit that nowhere does Loke disclose or suggest “the use of URLs with session identifiers which can be extracted when required,” as stated by the Examiner.

The Examiner also states, regarding Loke, that “Loke teaches the use of structured logic programming for various objectives in crawling a web site ... Loke also teaches the use of logic to compare URLs and attach state to URLs.” (Office Action, paragraph bridging pages 3 and 4). Regardless of the accuracy of these statements, Applicants fail to see how the Examiner’s discussion of Loke relates to the features recited in claim 1. Loke, as with DaCosta, simply does not disclose or suggest, as is recited in amended claim 1, “determining whether a web site corresponding to the URL uses session identifiers based on a

comparison of URLs that are within the document and that change between the at least two different copies of the document.”

For at least these reasons, Applicants submit that neither DaCosta nor Loke, either alone or in combination, discloses or suggests each of the features of independent claim 1, and accordingly, the rejection of this claim under 35 U.S.C. § 103(a) should be withdrawn. The rejection of claims 2 and 9 based on DaCosta and Loke should also be withdrawn, at least by virtue of the dependency of these claims from claim 1.

Dependent claims 2 and 9 recite additional features that are not disclosed or suggested by DaCosta and Loke. Amended claim 9, for example, recites that the comparison determines that the web site uses session identifiers when a portion of the URLs that change between the at least two different copies of the document is greater than a predetermined value. Neither DaCosta nor Loke in any way disclose or suggest this aspect of claim 9.

Independent claim 10 and its dependent claims 11-14 also stand rejected under 35 U.S.C. § 103(a) based on DaCosta and Loke. Applicants respectfully traverse this rejection.

Claim 10 is directed to a method for identifying web sites that use session identifiers. The method includes downloading at least two different copies of at least one document from a web site; extracting uniform resource locators (URLs) from the two different copies of the web document; comparing the extracted URLs of the two different copies of the document; and determining whether the web site uses session identifiers based on the comparison.

In rejecting claim 10, the Examiner relies on DaCosta to disclose “storing and identifying session identifiers URLs” and “the analysis of URLs and headers to determine if a web site uses session IDs.” (Office Action, page 5). Applicants once again respectfully disagree with the Examiner’s interpretation of DaCosta. As discussed above with respect to claim 1, DaCosta does not mention the use of session identifiers. Thus, DaCosta could not possibly disclose determining whether a web site uses session identifiers based on a comparison, as recited in claim 10.

Claim 10 also recites downloading at least two different copies of at least one document from a web site; extracting uniform resource locators (URLs) from the two different copies of the web document; and comparing the extracted URLs of the two different copies of the document. The Examiner appears to concede that DaCosta does not disclose or suggest many of these features, (Office Action, pages 5 and 6), but contends that these features would have been obvious modifications in view of Loke’s teaching of “structured logic programming for various objectives in crawling a web site.” (Office Action, page 6).

Applicants strongly disagree with the Examiner’s determination of obviousness based on DaCosta and Loke. Applicants concede that it was known in the art that web crawling could use web spiders designed using “structured logic programming.” Just because a web spider could be designed using a structured programming language, however, in no way suggests the specific acts recited in claim 1. The Examiner’s conclusion of obviousness is conclusory and appears to be entirely based on hindsight taken from Applicants’ own

specification. The Examiner has not made a proper *prima facie* case of obviousness under 35 U.S.C. § 103(a).

For at least these reasons, Applicants submit that DaCosta and Loke, either alone or in combination, do not disclose or suggest each of the features recited in claim 10. Accordingly, the rejection of claim 10 under 35 U.S.C. § 103(a) is improper and should be withdrawn. The rejections of claims 11-14 are also improper, at least by virtue of their dependency from claim 10.

Independent claim 15 and its dependent claims 16-20 also stand rejected under 35 U.S.C. § 103(a) based on DaCosta and Loke. Applicants respectfully traverse this rejection.

Claim 15 is directed to a device including a spider component configured to crawl web documents associated with at least one web site. The device additionally includes a session identifier component configured to determine whether the web site uses session identifiers based on a comparison of a portion of URLs that change between different copies of at least one web document downloaded from the web site.

As discussed above with respect to claim 1, DaCosta does not mention the use of session identifiers. Thus, DaCosta could not possibly disclose or suggest the session identifier component recited in claim 15, which determines whether the web site uses session identifiers based on a comparison of a portion of URLs that change between different copies of at least one web document downloaded from the web site.

In rejecting claim 15, the Examiner additionally relies on Loke to teach “the use of structured logic programming for various objectives in crawling a web site

... Loke also teaches the use of logic to compare URLs and attach state to URLs (p. 239, "Using the Notion of State)." (Office Action, page 8). Again, Applicants submit that the Examiner fails to make a proper *prima facie* case of obviousness. Loke's disclosure of a structured programming language in no way suggests using the programming language to create the session identifier component recited in claim 15. Additionally, Applicants disagree with the Examiner's statement that Loke discloses attaching state to URLs. The cited section of Loke clearly discusses viewing a web page as an object with state, not the URL associated with the web page. Applicants submit that Loke simply fails to in anyway disclose or suggest modifying DaCosta to include the session identifier component recited in claim 15. Thus, the Examiner has not made a proper *prima facie* case of obviousness under 35 U.S.C. § 103(a).

For at least these reasons, Applicants submit that DaCosta and Loke, either alone or in combination, do not disclose or suggest each of the features recited in claim 15. Accordingly, the rejection of claim 15 under 35 U.S.C. § 103(a) is improper and should be withdrawn. The rejections of claims 16-20 are also improper, at least by virtue of their dependency from claim 15.

Independent claim 21 and its dependent claims 22-25 also stand rejected under 35 U.S.C. § 103(a) based on DaCosta and Loke. Applicants respectfully traverse this rejection.

Claim 21 recites a number of features similar, although not identical in scope, to those recited in claim 10. Accordingly, based on rationale similar to that given for claim 10, Applicants submit that the rejection of claim 21 is

improper and should be withdrawn. The rejections of claims 22-25 are also improper, at least by virtue of their dependency from claim 21.

Independent claim 26 and its dependent claims 27-30 also stand rejected under 35 U.S.C. § 103(a) based on DaCosta and Loke. Applicants respectfully traverse this rejection.

Independent claim 26 recites a number of features similar, although not identical in scope, to those recited in claim 10. Accordingly, based on rationale similar to that given for claim 10, Applicants submit that the rejection of claim 26 is improper and should be withdrawn. The rejections of claims 27-30 are also improper, at least by virtue of their dependency from claim 26.

Independent claim 31 recites certain features similar, although not identical in scope, to those recited in claim 15. Accordingly, based on similar rationale, Applicants submit that the rejection of claim 31 is improper and should be withdrawn.

*Rejections Under 35 U.S.C.
§ 103(a) Based on DaCosta, Loke, and Ultraseek*

In rejecting dependent claims 3-8, the Examiner relies on Ultraseek in addition to DaCosta and Loke. More specifically, the Examiner appears to rely on Ultraseek for the disclosure of extracting a session identifier from a URL. (Office Action, page 12).

As an initial matter, Applicants note that Ultraseek is not prior art to the instant application. The Ultraseek document appears to state that it was created in January of 2001 and last updated in November of 2004. The instant application was filed on September 29, 2003. It is not clear how many times the

Ultraseek document was updated between September 29, 2003 and November of 2004, or what was changed in the updates. Accordingly, Applicants submit that Ultraseek is not prior art under 35 U.S.C. § 103(a) and the rejection of claims 3-8 should be withdrawn for at least this reason.

Additionally, Applicants submit that even if, for the sake of argument, Ultraseek was considered to be prior art to the instant Application, Ultraseek would still not cure the above-mentioned deficiencies of DaCosta and Loke with respect to claim 1. For example, although Ultraseek appears to disclose removing session identifiers from URLs, the removal appears to be based on a user enterable "regular expression" that defines how a particular session identifier is embedded within a URL. In other words, removing session identifiers, as disclosed by Ultraseek, assumes that the user already knows that the web site uses session identifiers and knows the "regular expression" that defines how the session identifiers are embedded within the URL. This in no way discloses or suggests, for example, as is recited in amended claim 1, from which claims 3-8 depend, determining whether a web site corresponding to the URL uses session identifiers based on a comparison of URLs that are within the document and that change between the at least two different copies of the document.

For at least these reasons, Applicants submit that the rejections of claims 3-8 under 35 U.S.C. § 103(a) are improper and should be withdrawn.


Conclusion

In view of the foregoing amendments and remarks, Applicants respectfully request the Examiner's reconsideration of this application, and the timely allowance of the pending claims.

To the extent necessary, a petition for an extension of time under 37 CFR 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account No. 50-1070 and please credit any excess fees to such deposit account.

Respectfully submitted,

HARRITY SNYDER, L.L.P.

By: 
Brian E. Ledell
Reg. No. 42,784

11350 Random Hills Road
Suite 600
Fairfax, Virginia 22030
(571) 432-0800

Date: February 23, 2006

Attachment: Replacement Drawing Sheet Fig. 8